DOCUMENT RESUME

ED 116 162 CS 002 326

AUTHOR	Hubert, Lawrence J.; Levin, Joel R.
TITLE	A General Statistical Framework for Assessing Categorical Clustering in Free Recall.
INSTITUTION	Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	WRDCCL-TP-58
PUB DATE	Oct 75
CONTRACT	NE-C-00-3-0065
NOTE	27p.
EDRS PRICE	MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS	*Cluster Analysis; *Cognitive Processes; Educational Research; Evaluation Methods; Higher Education; Measurement Instruments; *Recall (Psychological); Retention; *Statistical Analysis
IDENTIFIERS	*Free Recall

ABSTRACT

A graph-theoretic paradigm is used to generalize the common measures of categorical clustering in free recall based on the number of observed repetitions. Two graphs are defined: a graph G that characterizes the a priori structure of the item set defined by a researcher, and a graph R that characterizes a subject's protocol. Two indices of clustering, denoted by gamma and omega, are obtained by evaluating the sum of the pairwise products of the weights on the corresponding edges of the two graphs. The gamma statistic is a direct generalization of the commonly used clustering indices and reduces to the number of repetitions whenever G represents a standard categorical decomposition of a stimulus list. The omega statistic, on the other hand, extracts more information from the protocol graph R than does gamma and incorporates a distance measured based on the number of intervening items in a subject's recall sequence. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED116162

Theoretical Paper No. 58

A GENERAL STATISTICAL FRAMEWORK FOR ASSESSING CATEGORICAL CLUSTERING IN FREE RECALL

by

Lawrence J. Hubert and Joel R. Levin

Report from the Project on
Children's Learning and Development

Joel R. Levin
Principal Investigator

2

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

October 1975

CS002326

The paper was prepared while the first author was supported in part by a Postdoctoral Research Training Fellowship from the Social Science Research Council, and while the second author was a Visiting Fellow at the National Institute of Education.

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that agency should be inferred.

Center Contract No. NE-C-00-3-0065

WISCONSIN RESEARCH AND DEVELOPMENT CENTER FOR COGNITIVE LEARNING

MISSION

The mission of the Wisconsin Research and Development Center for Cognitive Learning is to help learners develop as rapidly and effectively as possible their potential as human beings and as contributing members of society. The R&D Center is striving to fulfill this goal by

- conducting research to discover more about how children learn
- developing improved instructional strategies, processes and materials for school administrators, teachers, and children, and
- offering assistance to educators and citizens which will help transfer the outcomes of research and development into practice

PROGRAM

The activities of the Wisconsin R&D Center are organized around one unifying theme, Individually Guided Education.

FUNDING

The Wisconsin R&D Center is supported with funds from the National Institute of Education; the Bureau of Education for the Handicapped, U.S. Office of Education; and the University of Wisconsin.

TABLE OF CONTENTS

Chapter		Page
	List of Tables	vii
	List of Figures	vii
	Abstract	ix
I.	Introduction	1
II.	A Graph-Theoretic Paradigm	3
III.	A Permutation Distribution for Γ	7
IV.	Mean and Variance for Γ	11
V.	An Alternative Index Ω and Some Extensions	13
VI.	Indices of Clustering	17
VII.	Group Statistics	21
VIII.	Discussion	25
	References	27

LIST OF TABLES

Table		Page
1	A Sample Permutation Distribution for Γ	8
2	A Sample Permutation Distribution for Ω	14
3	Possible Indices of Clustering: 0 Represents Either Γ or Ω	19

LIST OF FIGURES

Figure		Page
1	Illustration of a graph G on five nodes with non-negative weights attached to all edges (upper portion uses a general weight function; lower portion is a standard interpretation)	4

I

INTRODUCTION

The general problem of defining indices of categorical clustering in free recall has been the focus of extensive research in recent years (for instance, see Dalrymple-Alford, 1970; Frankel & Cole, 1971; Kelly, 1973; Roenker, Thompson, & Brown, 1971; and Shuell, 1969). Most of these contributions discuss alternative statistics that measure the degree to which a series of responses provided by a subject conforms to a hypothesized structure within the set consisting of all potential responses. Typically, a set of words or other stimuli that are assumed to be categorized into mutually exclusive and exhaustive classes is given to a subject to study in a randomized order; subsequently, the subject is asked to recall as many items as possible from memory. An index of clustering quantifies the amount of correspondence between the subject's protocol and the specific partition of the items hypothesized by the researcher. If clustering in recall occurs according to expectations, then the responses of a subject should be grouped more or less consistently with respect to the a priori categories that theoretically partition the original stimulus list, and in particular, there should be a tendency for related items to be recalled together.

The intent of this paper is <u>not</u> to propose yet another clustering index as a competitor to the numerous ones already "on the market" (for illustrations, refer to the papers cited earlier). Instead, we wish to provide a novel framework within which several of the more popular clustering indices may be viewed. In the first sections below, a graph-theoretic characterization of the clustering problem is developed; in the later sections certain specializations of the general framework are discussed along with the appropriate statistical inference procedures. As one further comment, it should be pointed out that the material to follow is limited to the categorical clustering problem rather than to free recall clustering in general (cf., Pellegrino's [1971] discussion of the subjective-organization paradigm).

II

A GRAPH-THEORETIC PARADIGM

As a convention, suppose S denotes the set of n stimuli $\{o_1,\ldots,o_n\}$ that contains the items presented to a subject. To formalize the underlying structure of the stimulus set, it is convenient to define a graph G that has n nodes or points $o_1,\ldots,o_n$ with an edge or line between each unordered pair of distinct nodes. A nonnegative weight is attached to each edge, that for notational purposes will be referred to as $q(o_i,o_j)$, where o_i and o_j are two distinct arbitrary nodes in S and define a single edge. The upper portion of Figure 1 illustrates the type of pictorial representation that may be given for any graph G. In this example, n is 5 and the arbitrary weights for all ten edges are between 0.0 and 1.0, as might be represented by various numerical association norms.

As a special case, a graph G may be used to represent any categorization assumed for the set S defined by a partition of S into object classes containing $n_1,\ldots,n_k$ elements, where $n_1+\ldots+n_k = n$. Note that this case encompasses object classes and their associated elements defined either in a priori terms (experimenter-defined) or on the basis of subjects' idiosyncracies (subject-defined), with the latter exemplified by subjects sorting objects into subject-perceived categories (cf., Mandler, 1967). In the present context, both types of categorization are considered to characterize the stimulus structure graph G. For a pair of nodes within the same object class of the partition, the weight function is defined to be 1.0; conversely, any edge between two nodes from separate object classes is assigned a weight of 0.0. For example, the lower portion of Figure 1 shows how the graph G would appear if five objects (n = 5) belonged to two classes (k = 2) with three objects, o_1, o_2, and o_3, in one class ($n_1 = 3$) and two objects, o_4 and o_5, in the other ($n_2 = 2$). For convenience, this particular case will be called the standard interpretation, but clearly, a categorization defined, say, by overlapping subsets or by a more complex structure could be characterized in a similar way.

In a related manner the response sequence provided by a subject can be represented by a second graph R on the node set $\{o_1,\ldots,o_n\}$. For the graph R the weight attached to an edge is either 1.0 or 0.0, where a 1.0 signifies that the two nodes were recalled sequentially with no intervening elements. Without loss of generality, it is assumed that all elements of S are actually recalled, since otherwise the original set S could be redefined as those elements listed by a subject.[1] Thus, the graph R consists of a single

---

[1] We do not wish to contest here whether the proper basis for clustering is the unconditional or the conditional stimulus set (cf., Frender & Doubilet, 1974). The procedures to be described can be applied in either case.

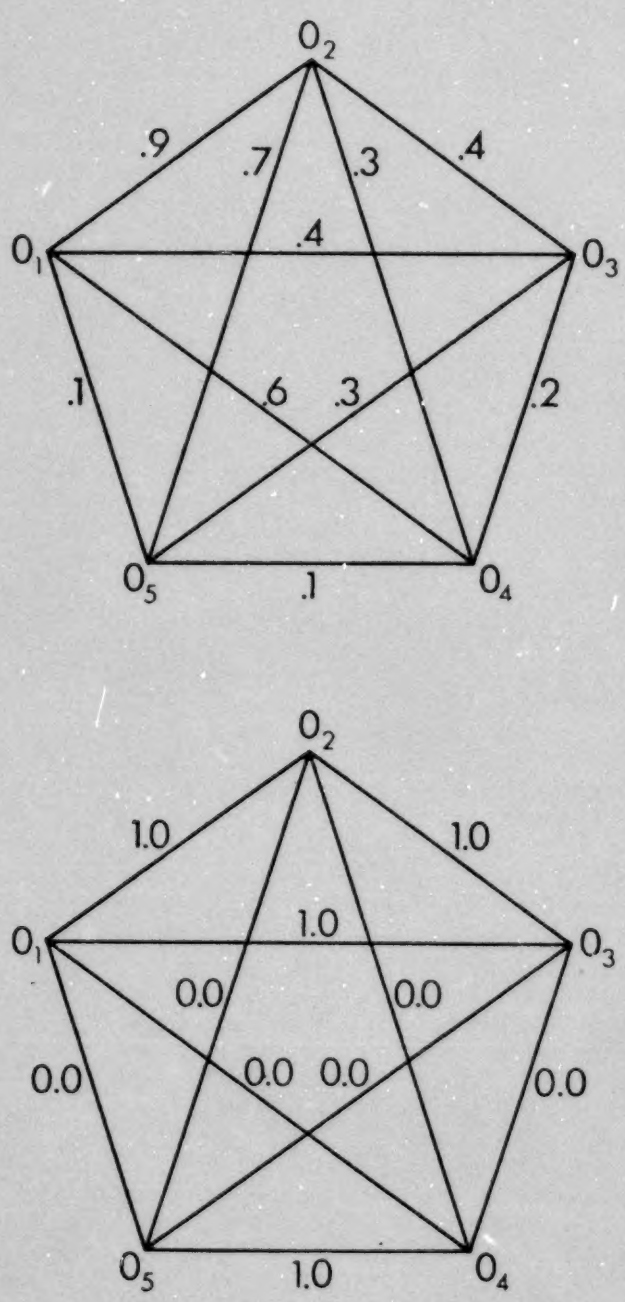

Figure 1. Illustration of a graph G on five nodes with nonnegative weights attached to all edges (upper portion uses a general weight function; lower portion is a standard interpretation).

contiguous sequence of edges all having weights of 1.0 that passes through each node once and only once.

One possible measure of correspondence between a subject's recall sequence and the hypothesized structure is given by the index Γ:

$$\Gamma = (1/2) \sum_{j=1}^{n} \sum_{i=1}^{n} q(o_i,o_j)C(o_i,o_j) = \sum_{i<j} \sum q(o_i,o_j)C(o_i,o_j),$$

where $C(o_i,o_j)$ is the zero-one weight function characterizing the graph R, $q(o_i,o_j)$ is as previously defined for G, and $q(o_i,o_i) = C(o_i, o_i) = 0$ for all i. In the standard interpretation, Γ is merely the number of repetitions, i.e., the number of node pairs that are recalled sequentially and belong to the same object class within the hypothesized partition. Since the number of repetitions or some transform of this quantity is the commonly used measure of clustering discussed in the literature, the Γ statistic is a natural generalization. Specifically, a large index Γ results when the node pairs that are recalled sequentially also have the larger associated weights on the defined edges in G. Although this discussion will emphasize the index Γ, an alternative measure will be proposed in a later section that incorporates more information from a subject's protocol than simple adjacent responses.

The constant multiplier of 1/2 used in the definition of Γ implies in an intuitive sense that some type of correction is being made for counting the same products twice. In particular, if the original index Γ were stated without the constant multiplier and the weight functions were not assumed to be symmetric, then a similar index may be defined between two possibly asymmetric weight functions. The graph G would be characterized by the presence of <u>two</u> edges between each pair of nodes o_i and o_j, where one edge is directed from o_i to o_j and weighted by $q(o_i,o_j)$ and the second edge is placed in an opposite orientation and weighted by $q(o_j,o_i)$, i.e., directed from o_j to o_i. In a similar way the protocol graph R could be directed; for instance, each edge that has a weight of 1.0 is matched with an edge between the same two nodes but with a weight of 0.0 and is directed in the opposite way. General directed graphs of this type could prove a useful extension if the order in which the subject provides the recalled nodes is of interest (e.g., see Pellegrino, 1971), but for our purposes only symmetric weight functions will be considered explicitly.

At this point there are two distinct problems that could be attacked: (a) normalizing the index Γ to provide a <u>measure</u> of clustering, or (b) defining a hypothesis-testing procedure for evaluating the size of an observed Γ. In some instances the second problem subsumes the first, since many of the more acceptable normalizations require an initial calculation of statistics that are also needed in hypothesis testing. Nevertheless, a number of possible normalizations will be presented later that relate directly to several of the more popular indices already used for the standard interpretation.

III

A PERMUTATION DISTRIBUTION FOR Γ

One possible strategy for assessing the correspondence between the two graphs R and G is to develop a statistical baseline through a randomization or permutation distribution for the index Γ (for example, see Barton & David, 1966). Under the assumption that there is no inherent relationship between a subject's response protocol and the underlying assumed structure defined by G, each possible permutation of the nodes $o_1,\ldots,o_n$ is assumed to have an equally likely chance of occurring a priori as the subject's response sequence. Since there are n! possible orderings of the n nodes, an index Γ could be calculated for each such sequence, generating what is typically called a permutation distribution for Γ. By comparing the observed value of Γ to this distribution, a precise evaluation may be made as to whether the observed value of Γ is large enough to reject the hypothesis that the subject's protocol has no inherent relationship to the researcher's theoretical categorization. In other words, with respect to the graph G the following question is raised: Is it reasonable to infer that the subject's protocol was not chosen at random from the n! possible response sequences?

Clearly there are many difficulties with this formulation, since even in the event that a subject is responding independently of the assumed categorization, it is very unlikely that the protocol chosen can be viewed realistically as an actual random selection from all n! possible response sequences formed from the list of recalled nodes (for example, see Shuell, 1969). Nevertheless, an inference technique based upon complete randomization is justified to the extent that response biases, such as serial position effects, are unrelated to the categorization being tested by the researcher.[2] There does not appear to be any simple way of making this obviously vague generalization any more precise that would, at the same time, allow the development of a very general inference procedure.

As a very elementary example that should provide some clarification, suppose that a subject recalls four words in the order o_1,o_2,o_3,o_4. The researcher has assumed that a standard interpretation holds in which the nodes $\{o_1,o_2\}$ form one category and $\{o_3,o_4\}$ form a second. In this illustration, two edges are present in G with weights of 1.0 between o_1 and o_2 and between o_3 and o_4; alternatively, in R, three edges are present with weights of 1.0 defined between each pair of adjacent responses: o_1 and o_2, o_2 and o_3, and o_3 and o_4. All other edges in both graphs have weights of 0.0. Consequently, the observed value of Γ is 2.0, and the appropriate

---

[2] Response biases of this kind (that act to disturb the nominal probability levels under the assumption of "equally likely" sequences) may be counteracted to some extent by the investigator, through such techniques as block randomization of items representing different categories, the inclusion of "buffer" items in the first and last few study list positions, and the insertion of an interpolated-activity interval between study and test. More complex decision rules could also be devised, such as ignoring those items in the subject's protocol that occur in exactly the same k initial or terminal serial positions as on the study list.

permutation distribution is defined by calculating Γ for all 4! = 24 possible response protocols, where each such protocol includes all four of the nodes (see Table 1).

TABLE 1

A SAMPLE PERMUTATION DISTRIBUTION FOR Γ

	Permutation	Γ Value
1-2:	$o_1\ o_2\ o_3\ o_4;\ o_4\ o_3\ o_2\ o_1$	2
3-4:	$o_1\ o_2\ o_4\ o_3;\ o_3\ o_4\ o_2\ o_1$	2
5-6:	$o_1\ o_3\ o_2\ o_4;\ o_4\ o_2\ o_3\ o_1$	0
7-8:	$o_1\ o_3\ o_4\ o_2;\ o_2\ o_4\ o_3\ o_1$	1
9-10:	$o_1\ o_4\ o_2\ o_3;\ o_3\ o_2\ o_4\ o_1$	0
11-12:	$o_1\ o_4\ o_3\ o_2;\ o_2\ o_3\ o_4\ o_1$	1
13-14:	$o_2\ o_1\ o_3\ o_4;\ o_4\ o_3\ o_1\ o_2$	2
15-16:	$o_2\ o_1\ o_4\ o_3;\ o_3\ o_4\ o_1\ o_2$	2
17-18:	$o_2\ o_3\ o_1\ o_4;\ o_4\ o_1\ o_3\ o_2$	0
19-20:	$o_2\ o_4\ o_1\ o_3;\ o_3\ o_1\ o_4\ o_2$	0
21-22:	$o_3\ o_1\ o_2\ o_4;\ o_4\ o_2\ o_1\ o_3$	1
23-24:	$o_3\ o_2\ o_1\ o_4;\ o_4\ o_1\ o_2\ o_3$	1

The probability distribution based on these obtained values of Γ is as follows:

Γ	Probability
0	8/24
1	8/24
2	8/24

Within a hypothesis-testing context, the probability of observing a value of Γ equal to 2 (or larger) is 1/3 under the assumption that the response protocol is chosen at random from the 4! possible sequences. A larger

value of n would be necessary to provide attainable significance levels in the traditional ranges of .05 to .01, but obviously, the same paradigm could be used with a corresponding increase in the required computational labor.

The procedure just described constructs what is called a "conditional permutation distribution" in the statistical literature, where the term "conditional" refers to the use of the subject's actual protocol in identifying a subset of nodes for the construction of the reference distribution. Inference procedures based upon these ideas form the basis for much of nonparametric statistics, and in fact, some of the same problems that appear in applying nonparametric techniques also cause difficulties in the free recall framework as well. Specifically, since the permutation distribution must be generated anew for each particular application, alternative approaches that bypass complete enumeration must be found. Generally, two different solutions are attempted in the statistics literature: the substitution of "scores" (for instance, ranks or normal deviates) for the original numerical observations that will allow a tabling of the permutation distribution that suffices for all applications; or secondly, deriving the mean and variance formulas for the appropriate test statistic and relying on large sample distributions for hypothesis testing.

Unfortunately, because of the great variability in the types of categorization structure, the latter alternative is the only possibility that can be entertained for the free recall problem. Consequently, the next task is to derive the mean and variance parameters for T. For an attempt to obtain complete probability distributions in the case of a standard interpretation, the reader should consult Kelly (1973).

IV

MEAN AND VARIANCE FOR Γ

The mean and variance parameters for Γ are easily derived and, surprisingly, are a special case of a much more general set of expressions given by Mantel (1967) in the biometrics literature. For convenience, suppose A_1, A_2, and A_3 are defined as follows:

$$A_1 = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} q(o_i,o_j) \right)^2;$$

$$A_2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} q(o_i,o_j) \right)^2;$$

$$A_3 = \sum_{i=1}^{n} \sum_{j=1}^{n} q(o_i,o_j)^2.$$

Then, using this notation,

$$E(\Gamma) = (1/n) \sum_{i=1}^{n} \sum_{j=1}^{n} q(o_i,o_j); \qquad [1]$$

$$Var(\Gamma) = (1/(n(n-1)))(A_1 - 2A_2) - (1/n^2)A_1 + (1/n)A_3. \qquad [2]$$

For the standard interpretation in which the assumed partition consists of object classes of sizes $n_1,\ldots,n_k$, formulas [1] and [2] reduce considerably to the forms given in [3] and [4], respectively:

$$E(\Gamma) = (1/n)\left(\sum_{i=1}^{n} n_i^2 \right) - 1; \qquad [3]$$

$$\mathrm{Var}(\Gamma) = (1/(n^2(n-1)))\left(\sum_{i=1}^{n} n_i^2\right)^2 \qquad [4]$$

$$- (2/(n(n-1)))\left(\sum_{i=1}^{n} n_i^3\right)$$

$$+ ((n+1)/(n(n-1)))\left(\sum_{i=1}^{n} n_i^2\right)$$

$$- n/(n-1).$$

In this special case, [3] is the expected number of repetitions and is identical to the expression derived by Bousfield and Bousfield (1966). Furthermore, the variance term in [4] is equivalent to a formula used by Frankel and Cole (1971) and is equal to the variance of the number of runs in a multiple-type object context since the number of such runs is merely the complement of the number of repetitions. For the probability distribution given previously and using formulas [3] and [4], we find that $E(\Gamma) = 1$ and $\mathrm{Var}(\Gamma) = 2/3$. These values can be verified numerically by computing the mean and variance of Γ directly from the complete permutation distribution.

Because the mean and variance parameters for Γ are available, it is natural to normalize the index Γ in the following way:[3]

$$Z = (\Gamma - E(\Gamma))/\sqrt{\mathrm{Var}(\Gamma)}.$$

This normalization "corrects" the observed value of Γ for the amount of clustering expected for the particular items recalled by the subject. Following Frankel and Cole (1971), the statistic Z generalizes the type of deviation measure that Shuell (1969) suggests for an index of clustering in a standard interpretation. Several other indices are suggested later. Finally, it should be noted that it seems reasonable to compare this Z index to a standard normal distribution (given relatively large n) in order to provide an approximation to the permutation distribution discussed earlier.

---

[3] Although in this section we discuss normalization procedures for a single subject under a conditional permutation model, a more useful extension can be developed through an appropriate measure $C(o_i, o_j)$ based on N protocols. This is presented in a later section.

V

AN ALTERNATIVE INDEX Ω AND SOME EXTENSIONS

The common measures of clustering used in the free recall literature, including the general measure Γ, depend only upon a minimal amount of information from an individual subject's protocol. Specifically, only those node pairs that are recalled sequentially contribute to the measure and all other pairs contribute nothing, even those that are separated by only one intervening node within the recall sequence. There is one rather simple scheme, however, for incorporating additional information from the subject's protocol by defining an alternative index Ω. Suppose the subject generates the node sequence $o_1,\ldots,o_n$ and a proximity function is defined between any two nodes in the protocol as the number of intervening nodes plus one. Thus, two nodes that are recalled far apart should have a large associated proximity function. In particular, define $C(o_r,o_s) = |r - s|$ and let

$$\Omega = (1/2) \sum_{i=1}^{n} \sum_{j=1}^{n} q(o_i,o_j)C(o_i,o_j)$$

$$= (1/2) \sum_{i=1}^{n} \sum_{j=1}^{n} q(o_i,o_j)|i - j| = \sum_{i<j} \sum q(o_i,o_j)(j-i).$$

If clustering in recall occurs, then two items o_i and o_j within the same category, or more generally, two items with relatively large values of $q(o_i,o_j)$, should have <u>small</u> associated function values $C(o_i,o_j)$. Consequently, the smaller the value of Ω, the more clustering in free recall occurs according to what is expected considering the weighted graph G. Fortunately, the mean and variance parameters for Ω are also available as special cases of the Mantel (1967) formulas:

$$E(\Omega) = [(n+1)/6] \sum_{i=1}^{n} \sum_{j=1}^{n} q(o_i,o_j); \qquad [5]$$

$$\mathrm{Var}(\Omega) = [(n+1)/180][-A_1+(n-4)A_2+4(n-1)A_3]. \qquad [6]$$

For the standard interpretation, these two expressions take on the simpler forms given in [7] and [8]:

$$E(\Omega) = ((n+1)/6)\left(\sum_{i=1}^{n} n_i^2 - n\right); \qquad [7]$$

$$\mathrm{Var}(\Omega) = ((n+1)/180)\left[-\left(\sum_{i=1}^{n} n_i^2\right)^2 + 4(n+1)\sum_{i=1}^{n} n_i^2 + (n-4)\sum_{i=1}^{n} n_i^3 - 4n^2\right]. \qquad [8]$$

As a simple numerical illustration in the case of a standard interpretation, the four-node example given previously may also be used to verify formulas [7] and [8]. In this case, the complete set of permutation values would be as shown in Table 2.

TABLE 2

A SAMPLE PERMUTATION DISTRIBUTION FOR Ω

	Permutation	Ω Value
1-2:	$o_1\ o_2\ o_3\ o_4;\ o_4\ o_3\ o_2\ o_1$	2
3-4:	$o_1\ o_2\ o_4\ o_3;\ o_3\ o_4\ o_2\ o_1$	2
5-6:	$o_1\ o_3\ o_2\ o_4;\ o_4\ o_2\ o_3\ o_1$	4
7-8:	$o_1\ o_3\ o_4\ o_2;\ o_2\ o_4\ o_3\ o_1$	4
9-10:	$o_1\ o_4\ o_2\ o_3;\ o_3\ o_2\ o_4\ o_1$	4
11-12:	$o_1\ o_4\ o_3\ o_2;\ o_2\ o_3\ o_4\ o_1$	4
13-14:	$o_2\ o_1\ o_3\ o_4;\ o_4\ o_3\ o_1\ o_2$	2
15-16:	$o_2\ o_1\ o_4\ o_3;\ o_3\ o_4\ o_1\ o_2$	2
17-18:	$o_2\ o_3\ o_1\ o_4;\ o_4\ o_1\ o_3\ o_2$	4
19-20:	$o_2\ o_4\ o_1\ o_3;\ o_3\ o_1\ o_4\ o_2$	4
21-22:	$o_3\ o_1\ o_2\ o_4;\ o_4\ o_2\ o_1\ o_3$	4
23-24:	$o_3\ o_2\ o_1\ o_4;\ o_4\ o_1\ o_2\ o_3$	4

The corresponding probability distribution would be:

Ω	Probability
2	8/24
4	16/24

Thus, computing either from formulas [7] and [8] or from the actual permutation distribution, we find $E(\Omega) = 10/3$ and $Var(\Omega) = 8/9$. A normalization of the index Ω using the mean and variance may be useful for interpretation here as well.

VI

INDICES OF CLUSTERING

Although hypothesis testing can be approached through an application of a randomization distribution, a second rather distinct problem still remains in defining "good" indices of categorical clustering. Exactly the same difficulty occurs in measuring rank correlation using the number of rank order inversions as a criterion. Almost all of the suggested rank correlation measures rely on the same statistic (usually denoted by S) to test the null hypothesis of no population association (see Hays, 1973, p. 799). Nevertheless, at least five different normalizations of this basic S statistic have been suggested as a way of providing a final measure of rank correlation, e.g., Somers' asymmetrical γ's, Goodman-Kruskal's γ, tau_a, tau_b, and tau_c (Somers, 1962). Consequently, the basic statistic for the standard interpretation free recall problem defined by the number of repetitions seems to be the natural analogue of the S statistic of rank correlation; moreover, the desire to find an adequate index of clustering corresponds directly to the historical search for a good index of rank correlation.

In our general framework, the indices Γ and Ω play the role of basic statistics that could be normalized in various ways to provide a final index of clustering. Several normalizations are suggested in Table 3 that will reduce for the special case of a standard interpretation to the more familiar measures discussed in the psychological literature. No attempt will be made to evaluate the merits of each of these normalizations, and thus, the reader is urged to consult the sources that are cited for extensive critiques and theoretical justifications.

Each of the indices given in Table 3 depends upon a number of constants chosen from the following list:

$$E(\Gamma),\ E(\Omega),\ \text{Var}(\Gamma),\ \text{Var}(\Omega),\ \text{Max}(\Gamma),\ \text{Max}(\Omega),\ \text{Min}(\Gamma),\ \text{Max}(\Omega).$$

All of these quantities have been defined earlier except for the Min and Max parameters, and these latter bounds can be obtained by a simple ordering operation. In particular, if the $n(n-1)/2$ values of $q(o_i,o_j)$ are ordered from smallest to largest and the $n(n-1)/2$ values of $C(o_i,o_j)$ are also ordered from smallest to largest, then one-half of the sum of the pairwise products of the two entries in the same rank position defines the maximum value of the index. Similarly, if the $n(n-1)/2$ values of $C(o_i,o_j)$ are reordered oppositely from largest to smallest, then one-half of the sum of the pairwise products defines the minimum index value (Gilmore, 1962). If a fairly simple structure for the graph G can be identified (e.g., a standard interpretation) then a closed-form expression for the minimum and maximum

TABLE 3

POSSIBLE INDICES OF CLUSTERING: θ Represents Either Γ or Ω

Index Name	Formula	Reference
Adjusted Ratio of Clustering	$[\theta - E(\theta)]/[Max(\theta) - E(\theta)]$	Roenker et al., 1971
C Index	$[\theta - Min(\theta)]/[Max(\theta) - Min(\theta)]$	Dalrymple-Alford, 1970
D Index	$[\theta - E(\theta)]/[Max(\theta) - Min(\theta)]$	Dalrymple-Alford, 1970
Z-Score Index	$[\theta - E(\theta)]/\sqrt{Var(\theta)}$	Hudson & Dunn, 1969; Frankel & Cole, 1971; Kelly, 1973
Deviation Index	$[\theta - E(\theta)]$	Bousfield & Bousfield, 1966
Modified Ratio of Repetition	$\theta/Max(\theta)$	Bower, Lesgold, & Tieman, 1969
Fagan's Index	$[\theta - E(\theta)]/Max(\theta)$	Fagan, 1968
Ratio of Repetition	$\theta/(n-1)$	Bousfield, 1953

index values may be obtained directly. In general, however, any application that depends on a rather complex structure in the graph G will require a separate evaluation of the minimum and maximum index values through this type of ordering procedure.

VII

GROUP STATISTICS

The indices of clustering in free recall that have been discussed up to this point are limited to the protocols of a single subject. However, there is an immediate generalization of the basic randomization paradigm that provides a direct extension to group data, or for that matter, to repeated trials using the same subject. For instance, suppose the stimulus structure graph G is fixed but we obtain N protocols either from a group of N subjects or from the same subject over N trials. Each of the N protocols is defined by a subset of the set of nodes S that define the graph G, and a proximity measure is constructed in some way between each pair of nodes o_i and o_j in S. As an illustration, an overall proximity function $C(o_i,o_j)$ could be obtained by first constructing for each protocol k a proximity matrix $C_k(o_i,o_j)$ between all node pairs in S and then summing (and possibly averaging) the N individual proximity functions. For a specific example, the proximity function $C_k(o_i,o_j)$ for protocols k could be defined as

$$C_k(o_i,o_j) = \begin{cases} 1 & \text{if } o_i \text{ and } o_j \text{ are recalled consecutively} \\ & \text{in protocol k;} \\ 0 & \text{otherwise.} \end{cases}$$

In this case, if $C(o_i,o_j) = \Sigma C_k(o_i,o_j)$, then the overall proximity between o_i and o_j is the number of protocols in which o_i and o_j were recalled sequentially. Thus, with this interpretation, larger values of $C(o_i,o_j)$ correspond to the more similar objects. As an alternative possibility, suppose that protocol k contains n_k recalled items and we define:

$$C_k(o_i,o_j) = \begin{cases} |i - j| & \text{if both } o_i \text{ and } o_j \text{ are recalled in protocol} \\ & \text{k and with } |i - j| - 1 \text{ intervening nodes;} \\ n_k + 1 & \text{if either } o_i \text{ or } o_j \text{ is not present in protocol k.} \end{cases}$$

Using this definition and summing over all protocols, small values of $C(o_i,o_j)$ denote the more similar object pairs.

In any event, given the final proximity measure $C(o_i,o_j)$, a general index may be defined by, say, Λ:

$$\Lambda = (1/2) \sum_{j=1}^{n} \sum_{i=1}^{n} q(o_i,o_j)C(o_i,o_j)$$

$$= \sum_{i<j} \sum q(o_i,o_j)C(o_i,o_j).$$

Mantel's formulas immediately provide the randomization mean and variance for Λ:

$$\text{Let } A_1 = \left(\sum_{j=1}^{n} \sum_{i=1}^{n} q(o_i,o_j) \right)^2;$$

$$A_2 = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} q(o_i,o_j) \right)^2;$$

$$A_3 = \sum_{j=1}^{n} \sum_{i=1}^{n} q(o_i,o_j)^2;$$

$$B_1 = \left(\sum_{j=1}^{n} \sum_{i=1}^{n} C(o_i,o_j) \right)^2;$$

$$B_2 = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} C(o_i,o_j) \right)^2;$$

$$B_3 = \sum_{j=1}^{n} \sum_{i=1}^{n} C(o_i,o_j)^2;$$

Then

$$E(\Lambda) = [1/(2n(n-1))]\sqrt{A_1B_1};$$

$$\text{Var}(\Lambda) = -[1/(2n(n-1))]^2 A_1B_1$$

$$+ [1/(2n(n-1))]A_3B_3$$

$$+ [1/(n(n-1)(n-2))][A_2 - A_3][B_2 - B_3]$$

$$+ [1/(4(n(n-1)(n-2)(n-3))][A_1 - 4A_2 + 2A_3]$$

$$[B_1 - 4B_2 + 2B_3].$$

With these parameters, a normalized Z statistic may be defined in the usual way:

$$Z = (\Lambda - E(\Lambda))/\sqrt{Var(\Lambda)}$$

Once again, this Z statistic should provide a convenient large-sample approximation to the exact permutation test that the measures $q(o_i,o_j)$ and $C(o_i,o_j)$ are unrelated, or more simply, Z could be used as a normalized group measure of clustering in free recall.

Although the general statistic Λ may be used to index clustering in free recall for a group of subjects or for a single subject over trials, a more traditional approach to group analyses should be noted. Here the single protocol statistics, say Γ or Ω, are calculated and used in traditional analysis of variance paradigms to assess group differences, trends, and so on. Clearly, the use of a clustering index as a dependent variable is a much more general technique than the simple randomization extension defined through the single index Λ.

VIII

DISCUSSION

Although the inference problem discussed in this paper has been framed completely within the free recall paradigm, in actuality the task of comparing two graphs can be made much more general. We have indicated earlier that in the free recall paradigm, the subject response graph, R, is compared with the stimulus structure graph, G, with the latter defined either by the experimenter or by the subject. In some cases, however, the stimulus structure graph may be of interest in its own right, namely, when an investigator wishes to compare some a priori structure with the subject's perception of it (see, for example, Anglin, 1970). Suppose the subject is asked to sort the elements of S into groups of similar objects, as is done in the Mandler (1967) paradigm. An index of correspondence between the subject's sort and the a priori structure characterized by G can be obtained in the same way that Γ or Ω were defined earlier.

In summary, the problem of comparing two graphs R and G appears to be a very general inference technique that can be identified as basic to many experimental situations in the behavioral sciences. Given the elegance of the associated randomization procedures, this framework is capable of providing an extremely general inference strategy. The necessary correspondences are now being developed in detail by the authors, and hopefully, this work will provide the applied researcher with a new set of useful and powerful analytical tools.

REFERENCES

Anglin, J. M. The growth of word meaning. Cambridge, Mass.: The M.I.T. Press, 1970.

Barton, D. F. & David, F. N. The random intersection of two graphs. In F. N. David (Ed.), Research papers in statistics. New York: Wiley, 1966. Pp. 445-459.

Bousfield, A. K. & Bousfield, W. A. Measurement of clustering and sequential constancies in repeated free recall. Psychological Reports, 1966, 19, 935-942.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology, 1953, 49, 229-240.

Bower, G. H., Lesgold, A. M., & Tieman, D. Grouping operations in free recall. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 481-493.

Dalrymple-Alford, E. C. The measurement of clustering in free recall. Psychological Bulletin, 1970, 75, 32-34.

Fagan, J. F. Measuring verbal recall II: The ITR score expressed as a ratio. Psychonomic Science, 1968, 11, 205.

Frankel, F. & Cole, M. Measures of category clustering in free recall. Psychological Bulletin, 1971, 76, 39-44.

Frender, R. & Doubilet, P. More on measures of category clustering in free recall--although probably not the last word. Psychological Bulletin, 1974, 81, 64-66.

Gilmore, P. C. Optimal and suboptimal algorithms for the quadratic assignment problem. SIAM Journal of Applied Mathematics, 1962, 10, 305-313.

Hays, W. L. Statistics for psychologists. New York: Holt, Rinehart, & Winston, 1973.

Hudson, R. L., & Dunn, J. E. A major modification of the Bousfield (1966) measure of category clustering. Behavior Research Methods & Instrumentation, 1969, 1, 110-111.

Kelly, M. E. Application of the theory of combinatorial chance to the estimation of significance of clustering in free recall. The British Journal of Mathematical and Statistical Psychology, 1973, 26, 270-280.

Mandler, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), The psychology of learning and motivation. Vol. 1. New York: Academic Press, 1967. Pp. 327-372.

Mantel, N. The detection of disease clustering and a generalized regression approach. Cancer Research, 1967, 27, 209-220.

Pellegrino, J. W. A general measure of organization in free recall for variable unit size and interval sequential consistency. Behavioral Research Methods and Instrumentation, 1971, 3, 241-246.

Roenker, D. L., Thompson, C. P., & Brown, S. C. Comparison of measures for the estimation of clustering in free recall. Psychological Bulletin, 1971, 76, 45-48.

Shuell, T. J. Clustering and organization in free recall. Psychological Bulletin, 1969, 72, 353-374.

Somers, R. H. A new asymmetric measure of association for ordinal variables. American Sociological Review, 1962, 27, 799-811.